DOCUMENT RESUME

ED 427 043 TM 029 421

AUTHOR	Barnette, J. Jackson; McLean, James E.
TITLE	The Tukey Honestly Significant Difference Procedure and Its Control of the Type I Error-Rate.
PUB DATE	1998-11-04
NOTE	15p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (27th, New Orleans, LA, November 4-6, 1998).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Comparative Analysis; *Error of Measurement; Monte Carlo Methods; *Research Methodology; *Sample Size; Simulation
IDENTIFIERS	*Tukey Statistic; *Type I Errors

ABSTRACT

Tukey's Honestly Significant Difference (HSD) procedure (J. Tukey, 1953) is probably the most recommended and used procedure for controlling Type I error rate when making multiple pairwise comparisons as follow-ups to a significant omnibus F test. This study compared observed Type I errors with nominal alphas of 0.01, 0.05, and 0.10 compared for various sample sizes and numbers of groups. Monte Carlo methods were used to generate replications expected to provide 0.95 confidence intervals of +/- 0.001 around the nominal alphas of 0.10, 0.05, and 0.01 for 42 combinations of n (5, 10, 15, 20, 30, 60, and 100) and numbers of groups (3, 4, 5, 6, 8, and 10). Means and standard deviations of observed Type I error rates and percentages of observed Type I errors falling below, within, and above the 0.95 confidence intervals were determined for total number of Type I errors. The results indicate that HSD is conservative relative to experimentwise Type I error control across all alpha levels, sample sizes, and number of groups. However, when per-experiment (total Type I errors) is of interest, HSD was liberal at alpha of 0.10 and 0.05, but was very conservative when alpha was 0.01. Results also point out the differences inherent in selection of a Type I error mode of control. Differences between per-experiment and experimentwise Type I error control was mostly a function of the number of groups being compared. As the number of groups increased, the difference between per-experiment and experimentwise error proportions increased. However, sample size was also a significant predictor; as sample size increased, the difference decreased. (Contains 9 tables and 14 references.) (Author/SLD)

The Tukey Honestly Significant Difference Procedure

and Its Control of the Type I Error-Rate

J. Jackson Barnette
University of Iowa

and

James E. McLean
University of Alabama at Birmingham

Address correspondence to: Dr. Jack Barnette
College of Medicine
1-204 Medical Education Building
University of Iowa
Iowa City, IA 52242-1000
(319) 335 8905
jack-barnette@uiowa.edu

A paper presented at the 1998 Annual meeting of the
Mid-South Educational Research Association, New Orleans, LA
November 4, 1998

Abstract

Tukey's HSD is probably the most recommended and used procedure for controlling Type I error-rate when making multiple pairwise comparisons as follow-ups to a significant omnibus F test. The purpose of this study was to compare observed Type I errors with nominal alphas of .01, .05, and .10 for various sample sizes and numbers of groups. Monte Carlo methods were used to generate replications expected to provide .95 confidence intervals of +/- .001 around the nominal alphas of .10, .05, and .01 for 42 combinations of n (5, 10, 15, 20, 30, 60, and 100) and numbers of groups (3, 4, 5, 6, 8, and 10). Means and standard deviations of observed Type I error-rates and percentages of observed Type I errors falling below, within, and above the .95 CI's were determined for total number of Type I errors. The results indicate HSD is conservative relative to experimentwise Type I error control across all alpha levels, sample sizes, and number of groups. However when per-experiment (total Type I errors) is of interest, HSD was liberal at alpha of .10 and .05, but was very conservative when alpha was .01. Results also point out the differences inherent in selection of a Type I error mode of control. Differences between per experiment and experimentwise Type I error control was mostly a function of the number of groups being compared. As the number of groups increased, the difference between per-experiment and experimentwise error proportions increased. However, sample size was also a significant predictor; as sample size increased the difference decreased.

THE TUKEY HONESTLY SIGNIFICANT DIFFERENCE PROCEDURE AND ITS CONTROL OF THE TYPE I ERROR-RATE

Whenever a researcher has more than two comparisons to test, control of the Type I error-rate becomes a concern. Soon after Fisher developed the process of analysis of variance (ANOVA), he recognized the potential problem of the error-rate becoming inflated when multiple t-tests were performed on three or more groups. He discusses this problem in the 1935 edition of his famous book, *The Design of Experiments*. His recommendation of using a more stringent alpha when performing his Least Significant Difference Procedure (LSD) is based on this concern. However, researchers still criticized the LSD as providing inadequate control of Type I error. This early recognition of the problem has resulted in hundreds of multiple comparison procedures being developed over the years. One of the more popular of these procedures (Toothaker, 1993) has been the Honestly Significance Difference Procedure (HSD).

Most studies of Type I error rates for follow-up of pairwise mean differences have been based on what is referred to as experimentwise or familywise error control philosophies. These terms are credited to Miller (1966). Experimentwise (EW) Type I error relates to finding at least one significant difference by chance for the specified alpha level. In these cases, the only difference of concern is the largest mean difference. Experimentwise Type I error control ignores the possibility of multiple Type I errors in the same experiment. The pairwise mean differences for those other than the largest mean difference are not considered. Type I error control is such that not all possible Type I errors are evaluated. In these cases, many procedures such as Tukey's HSD are considered to have conservative Type I error control since the actual probabilities of finding at least one Type I error are lower than the nominal alpha level.

Per-experiment (PE) Type I error control considers all the possible Type I errors that can occur in a given experiment. Thus, more than one Type I error per experiment is possible and reasonably likely to occur if there is an experimentwise Type I error on the highest mean difference. Klockars and Hancock (1994) pointed out the importance and risks associated with this distinction. They found, using a Monte Carlo simulation, that there was a difference of .0132 in the per-experiment and experimentwise Type I error rates for Tukey's HSD when alpha was set at .05. This discussion was expanded in their 1996 review titled "The Quest for α" (Hancock & Klockars). Thus, when one has exact control of Type I error in the experimentwise situation, the per-experiment Type I error probability is higher. One of the purposes of this research was to examine how much of a difference there may be between experimentwise and per-experiment Type I error rates for Tukey's HSD when used with alpha levels of .10, .05, and .01 and to determine the relative influence on this difference of number of groups and number of subjects per group. It is predicted that as the number of groups increases, the difference between experimentwise and per-experiment error rates will increase since there are more pairwise differences in the set. While most Type I error research is based on an experimentwise mode, the per-experiment Type I error is more consistent with the reality of pairwise hypothesis testing. It is not only the largest mean difference subjected to error control, but all the pairwise differences. Thus, we favor a per-experiment mode of Type I error control.

Background

The earliest example of what we now know as a multiple comparison procedure could be found in 1929, when Working and Hotelling applied simultaneous confidence intervals to regression lines. The Fisher (1935) reference cited earlier was the first application to the process of ANOVA. The Type I error-rate control problem was also referred to by Pearson and Sekar in 1936 and Newman in 1939. Newman described a multiple comparison test that used the "Studentized Range Statistic." It is said that his work was prompted by a discussion he had with Student. Years later, Keuls published an updated version of the procedure (1952) using the Studentized range. We now know that multiple comparison procedure as the Student-Newman-Kuels Procedure.

Probably the most-used multiple comparison procedure is the Honestly Significance Difference (HSD) (Toothacker, 1993). Tukey's Honestly Significant Difference Procedure (HSD) was presented originally in an non-published paper by Tukey in 1953. Based on the Studentized range statistic originally derived by Gosset (a.k.a., Student) (1907-1938), the original derivation assumed normality, homogeneity of variance, and an equal number of observations per group. The Studentized range statistic, unlike the t-statistic, takes into account the number of means being compared, adjusting for the total number of tests to make all pairwise comparisons (Kennedy & Bush, 1985). Purportedly, the HSD gives a per-experiment error rate. Many common statistical texts recommend it be used as a protected test for all post-hoc tests (e.g., Hayes, 1988; Kennedy & Bush, 1985; Kirk, 1982; Maxwell & Delaney, 1990). Doing so, it is supposed to control the experimentwise error at a preselected value such as .05 or .01.

Later refinements of the HSD demonstrated its robustness to violations of the normality and homogeneity assumptions. Replacing the sample size of the groups in the formula with the harmonic mean of the sample sizes has been described (Kirk, 1982) as an alternative when the sample sizes are not equal. In any case, this study uses only cases where the sample sizes are equal among groups and the other assumptions are achieved.

Methodology

Monte Carlo methods were used to generate the data for this research. All data comprising the groups whose means were compared were generated from a random normal deviate routine, which was incorporated into a larger compiled QBASIC program that conducted all needed computations. The program was written by the senior author. All sampling and computation, conducted with double-precision, routines were verified using SAS® programs. The program was run on a Dell Pentium II, 266 MHZ personal computer.

Several sample size and number of groups arrangements were selected to give a range of low, moderate, and large case situations. The number of groups was: 3, 4, 5, 6, 8, and 10 and the sample sizes for each group were: 5, 10, 15, 20, 30, 60, and 100, which when crossed gave 42 experimental conditions. This was replicated for three nominal alphas of .10, .05, and .01. The approach used was to determine what number of replications would be needed to provide an expected .95 confidence interval of +/- .001 around the nominal alpha. This is an approach to examination of how well observed Type I error proportions are reasonable estimates of a standard nominal alpha. In other words, if alpha is the standard, what proportion of the estimates of actual Type I error proportions can be considered accurate, as evidenced by them being within the expected .95 confidence interval around nominal alpha?

This was based on the assumption that errors would be normally distributed around the binomial proportion represented by nominal alpha. Thus, when alpha was .10, 345742 replications were needed to have a .95 confidence interval of +/- .001 or between .099 and .101. When alpha was .05, 182475 replications were needed to have a .95 confidence interval of +/- .001 or between .049 and .051 and when alpha was .01, 38032 replications were needed to have a .95 confidence interval of +/- .001 or between .009 and .011. Observed Type I error proportions falling into the respective .95 confidence intervals are considered to be reliable estimates of the expected Type I error rate. Observed Type I error proportions falling below the .95 CI are considered to be conservative, and observed Type I error proportions falling above the .95 CI are considered to be liberal.

Within each nominal alpha/sample size/number of groups configuration, the number of ANOVA replications were generated. Each replication involved drawing of elements of the sample from a distribution of normal deviates, computation of sample means, and the omnibus F test. If the omnibus F test was significant, Tukey's HSD was computed and all pairwise mean differences were compared with the HSD critical value. While HSD uses only one critical value for all differences, the pairwise differences were recorded in a hierarchical fashion to determine pairwise differences significant at each of the numbers of steps between means from K down to 2. This approach permitted determination of experimentwise Type I error (at least one Type I error per experiment) or a Type I error for the largest mean difference, and per-experiment Type I errors or the total number of Type I errors observed regardless of where they are in the stepwise structure.

Summary statistics were computed for each alpha level for experimentwise and per-experiment conditions including: the mean proportion of Type I errors, standard deviation of the proportion of Type I errors, minimum proportion, maximum proportion, and the percentage of proportions falling in the three regions associated with the .95 confidence interval. Additional analysis included computation of differences between per-experiment proportions and experimentwise proportions (PE-EW) and correlation/regression analysis to determine relative influences of number of groups and sample sizes on this difference.

Results

The results were presented separately for alphas of .10, .05, and .01. In each case, the trends will be discussed and a conclusion stated.

Type I errors when alpha= .10

Tables 1 and 2 present the proportions and summary statistics of Type I errors when experimentwise and pre-experiment modes are used for the varying number of groups and sample sizes. The mean proportion of Type I errors for the experimentwise mode was .081348 and 100% of the 42 mean proportions were below the expected .95 confidence interval. Thus, in this case HSD was conservative to making Type I errors. HSD became more conservative relative to experimentwise Type I errors as the number of groups increased.

However, when per-experiment rate is considered, the mean proportion of Type I errors was .127948 and 100% of the observed proportions were above the expected .95 confidence interval, indicating a liberal condition. As the number of groups increased, HSD became more liberal. The difference between the per-experiment and experimentwise proportions, as indicated in Table 3, was .04660. The correlation of number of groups with difference was .8899 ($p < .0001$) and the correlation of sample size and difference was -.2213 ($p = .1590$). These were entered into a regression equation and the first variable entered (number of groups) accounted for 79.2% of the variance; the addition of the sample size variable increased the variance-accounted-for percentage to 84.1. Clearly, in this case, the number of groups was the major factor associated with this difference.

Type I errors when alpha= .05

Tables 4 and 5 present the proportions and summary statistics of Type I errors when experimentwise and pre-experiment modes are used for the varying number of groups and sample sizes. The mean proportion of Type I errors for the experimentwise mode was .038655 and 100% of the 42 mean proportions were below the expected .95 confidence interval. Thus, in this case HSD was conservative to making Type I errors and tended to become more conservative as the number of groups increased.

However, when per-experiment rate is considered, the mean proportion of Type I errors was .055310 and 97.6% of the observed proportions were above the expected .95 confidence interval, indicating a liberal condition. The tendency to become more liberal as the number of groups increased was not as apparent as it had been in the alpha= .10 condition. The difference between the per-experiment and experimentwise proportions, as indicated in Table 6, was .01665. The correlation of number of groups with difference was .8266 ($p < .0001$) and the correlation of sample size and difference was -.2833 ($p = .0690$). These were entered into a regression equation and the first variable entered (number of groups) accounted for 68.3% of the variance; the addition of the sample size variable increased the variance-accounted-for percentage to 76.4. Clearly, in this case, the number of groups was the major factor associated with this difference. However, the relative influence decreased for number of groups and increased for sample size when compared with the alpha equal to .10 condition.

Type I errors when alpha= .01

Tables 7 and 8 present the proportions and summary statistics of Type I errors when experimentwise and pre-experiment modes are used for the varying number of groups and sample sizes. The mean proportion of Type I errors for the experimentwise mode was .007019 and 97.6% of the 42 mean proportions were below the expected .95 confidence interval. Thus, in this case HSD was conservative to making Type I errors and tended to become more conservative as the number of groups increased.

However, when per-experiment rate is considered, the mean proportion of Type I errors was .008772 and 57.1% of the observed proportions were below the expected .95 confidence interval while 42.9% were within the .95 confidence interval. Conditions of small numbers of groups and/or small sample sizes were more likely to have proportions within the .95 confidence interval while conditions with higher numbers of groups and larger sample sizes tended to have proportions falling below the .95 confidence internal or being conservative.

The difference between the per-experiment and experimentwise proportions, as indicated in Table 9, was .001753. The correlation of number of groups with difference was .6588 ($p < .0001$) and the correlation of sample size and difference was -.4091 ($p =$ .0071). These were entered into a regression equation and the first variable entered (number of groups) accounted for 43.4% of the variance; the addition of the sample size variable increased the variance-accounted-for percentage to 60.1. Clearly, in this case, the number of groups was the major factor associated with this difference. However, the relative influence decreased for number of groups and increased for sample size when compared with the alpha equal to .10 and .05 conditions.

Conclusions

In general, Tukey's HSD does not provide for accurate Type I error control either when experimentwise or per-experiment is the control philosophy. Its accuracy is a function of alpha, number of groups, sample sizes, and control philosophy. As it has been used in most applications, it provides for conservative experimentwise control and liberal per-experiment control. Tukey's HSD is conservative if experimentwise Type I error is the philosophy regardless of alpha, number of groups, or sample sizes when conducted as a protected test.

If per-experiment is the Type I error philosophy, HSD is liberal if alpha is .10 or .05; but when alpha is .01, it is accurate when few groups or small samples are used and tends to become conservative when more and larger samples are used. The difference between experimentwise and per-experiment proportions of Type I errors is large enough to be a concern in having confidence that differences are not likely to be a function of a pre-specified level of chance, particularly when alpha is higher.

The most important determinate of the difference between experimentwise and pre-experiment Type I error proportions is the number of groups. As the number of groups increases, the difference increases. The influence of sample size is inverse: as sample size increases, the difference between experimentwise and pre-experiment Type I error proportion decreases. As alpha decreases, the influence of number of groups decreases while the influence of sample size increases.

Additional study of the discrepancy between experimentwise and per-experiment Type I errors is needed. We need to determine just how important this discrepancy is. The current study did not consider the case of unequal sample sizes or heterogenous variances. Is it the same under conditions of unequal sample sizes and/or variances?

The determination of factors relating to the discrepancy remains to be done. The current study only included the HSD procedure conducted as a protected test. Does this phenomenon occur with other MCPs? The determination of why these patterns change as alpha changes also remains to be done.

Ultimately, there remains a need to identify or develop a multiple comparison procedure that will provide true per-experiment Type I error control for examining pairwise differences under a number of conditions. This study has demonstrated that the Monte Carlo method might be an effective means of doing so.

References

Fisher, R. A. (1935, 1960). *The design of experiments*, 7th ed. London: Oliver & Boyd; New York: Hafner.

Hancock, G. R., & Klockars, A. J. (1996). The quest for α: Developments in multiple comparison procedures in the quarter century since Games (1971), *Review of Educational Research, 66*, 269-306.

Hayes, W. L. (1988). *Statistics* (4th ed). New York: Holt, Rinehart, and Winston, Inc.

Kennedy, J. J., & Bush, A. J. (1985). *An introduction to the design and analysis of experiments in behavioral research*. Lanham, MD: University Press of America, Inc.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. (2nd ed). Belmont, CA: Brooks Cole.

Klockars, A. J. & Hancock, G. R. (1994). Per-experiment error rates: The hidden costs of several multiple comparison procedures. *Educational and Psychological Measurement, 54* (2), 292-298.

Keuls, M. (1952). The use of "Studentized range" in connection with an analysis of variance. *Euphytica*, 1, 112-122.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth Publishing Company.

Miller, R. G., (1966). *Simultaneous statistical inference*. New York: McGraw-Hill.

Newman, D. (1939). The distribution of the range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika, 31*, 20-30.

Pearson, E. S., & Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, 28, 308-320.

Toothaker, L.E. (1993). *Multiple comparison procedures*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-089. Newbury Park, CA: Sage.

Tukey, J. W. (1953). *The problem of multiple comparisons*. Unpublished manuscript, Princeton University.

Working, H., & Hotelling, H. (1929). Application of the theory of error to the interpretation of trends. *Journal of the American Statistical Association, 35*, 73-85.

Table 1. Observed Experimentwise Type I Errors for HSD, Alpha= .10, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	.0932 ↓	.0924 ↓	.0925 ↓	.0923 ↓	.0916 ↓	.0918 ↓	.0917 ↓
4	.0882 ↓	.0879 ↓	.0877 ↓	.0874 ↓	.0870 ↓	.0866 ↓	.0873 ↓
5	.0849 ↓	.0840 ↓	.0835 ↓	.0839 ↓	.0829 ↓	.0829 ↓	.0827 ↓
6	.0814 ↓	.0801 ↓	.0802 ↓	.0797 ↓	.0798 ↓	.0794 ↓	.0797 ↓
8	.0770 ↓	.0753 ↓	.0742 ↓	.0740 ↓	.0733 ↓	.0734 ↓	.0743 ↓
10	.0727 ↓	.0712 ↓	.0703 ↓	.0702 ↓	.0689 ↓	.0689 ↓	.0702 ↓

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.081348	.007553	.0689	.0932	100	0	0

Table 2. Observed Per-Experiment (Total) Type I Errors for HSD, Alpha= .10, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	.1217 ↑	.1160 ↑	.1151 ↑	.1143 ↑	.1132 ↑	.1132 ↑	.1125 ↑
4	.1335 ↑	.1252 ↑	.1236 ↑	.1221 ↑	.1208 ↑	.1198 ↑	.1200 ↑
5	.1411 ↑	.1307 ↑	.1279 ↑	.1268 ↑	.1246 ↑	.1238 ↑	.1227 ↑
6	.1455 ↑	.1330 ↑	.1305 ↑	.1283 ↑	.1266 ↑	.1263 ↑	.1256 ↑
8	.1510 ↑	.1374 ↑	.1317 ↑	.1308 ↑	.1274 ↑	.1267 ↑	.1283 ↑
10	.1516 ↑	.1369 ↑	.1317 ↑	.1516 ↑	.1297 ↑	.1259 ↑	.1287 ↑

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.127948	.009794	.1125	.1516	0	0	100.0

Table 3. Differences in Per-Experiment and Experimentwise Type I Errors for HSD, Alpha= .10

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100	Mean
3	.0285	.0236	.0226	.0220	.0216	.0214	.0208	.0229
4	.0453	.0373	.0359	.0347	.0338	.0332	.0327	.0361
5	.0562	.0467	.0444	.0429	.0417	.0409	.0400	.0447
6	.0641	.0529	.0503	.0486	.0468	.0469	.0459	.0508
8	.0740	.0621	.0575	.0568	.0541	.0533	.0540	.0588
10	.0789	.0657	.0614	.0814	.0608	.0570	.0585	.0662
Mean	.0578	.0481	.0454	.0477	.0431	.0421	.0420	.0466

Table 4. Observed Experimentwise Type I Errors for HSD, Alpha= .05, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	.0461 ↓	.0455 ↓	.0455 ↓	.0452 ↓	.0453 ↓	.0446 ↓	.0451 ↓
4	.0434 ↓	.0422 ↓	.0419 ↓	.0421 ↓	.0427 ↓	.0418 ↓	.0422 ↓
5	.0404 ↓	.0401 ↓	.0398 ↓	.0399 ↓	.0389 ↓	.0397 ↓	.0401 ↓
6	.0394 ↓	.0379 ↓	.0380 ↓	.0371 ↓	.0368 ↓	.0378 ↓	.0380 ↓
8	.0364 ↓	.0345 ↓	.0339 ↓	.0338 ↓	.0339 ↓	.0336 ↓	.0347 ↓
10	.0337 ↓	.0324 ↓	.0321 ↓	.0317 ↓	.0319 ↓	.0313 ↓	.0321 ↓

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.038655	.004578	.0313	.0461	100	0	0

Table 5. Observed Per-Experiment (Total) Type I Errors for HSD, Alpha= .05, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	.0573 ↑	.0542 ↑	.0541 ↑	.0532 ↑	.0532 ↑	.0524 ↑	.0523 ↑
4	.0612 ↑	.0559 ↑	.0548 ↑	.0547 ↑	.0548 ↑	.0536 ↑	.0538 ↑
5	.0621 ↑	.0573 ↑	.0565 ↑	.0554 ↑	.0537 ↑	.0542 ↑	.0544 ↑
6	.0640 ↑	.0574 ↑	.0562 ↑	.0543 ↑	.0532 ↑	.0546 ↑	.0544 ↑
8	.0638 ↑	.0573 ↑	.0551 ↑	.0539 ↑	.0530 ↑	.0522 ↑	.0533 ↑
10	.0635 ↑	.0557 ↑	.0533 ↑	.0533 ↑	.0528 ↑	**.0504**	.0522 ↑

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.055310	.003240	.0504	.0640	0	2.4	97.6

Table 6. Differences in Per-Experiment and Experimentwise Type I Errors for HSD, Alpha= .05

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100	Mean
3	.0112	.0087	.0086	.0080	.0079	.0078	.0072	.0085
4	.0178	.0137	.0129	.0126	.0121	.0118	.0116	.0132
5	.0217	.0172	.0167	.0155	.0148	.0145	.0143	.0164
6	.0246	.0195	.0182	.0172	.0164	.0168	.0164	.0184
8	.0274	.0228	.0212	.0201	.0191	.0186	.0186	.0211
10	.0298	.0233	.0212	.0216	.0209	.0191	.0201	.0223
Mean	.0221	.0175	.0165	.0158	.0152	.0148	.0147	.0164

Table 7. Observed Experimentwise Type I Errors for HSD, Alpha= .01, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	.0089 ↓	.0085 ↓	**.0092**	.0080 ↓	.0086 ↓	.0086 ↓	.0085 ↓
4	.0079 ↓	.0083 ↓	.0079 ↓	.0085 ↓	.0084 ↓	.0076 ↓	.0075 ↓
5	.0077 ↓	.0069 ↓	.0067 ↓	.0073 ↓	.0071 ↓	.0073 ↓	.0074 ↓
6	.0074 ↓	.0070 ↓	.0072 ↓	.0064 ↓	.0069 ↓	.0062 ↓	.0070 ↓
8	.0065 ↓	.0059 ↓	.0060 ↓	.0060 ↓	.0060 ↓	.0057 ↓	.0060 ↓
10	.0059 ↓	.0053 ↓	.0055 ↓	.0055 ↓	.0050 ↓	.0052 ↓	.0054 ↓

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.007019	.001157	.0050	.0092	97.6	2.4	0

Table 8. Observed Per-Experiment (Total) Type I Errors for HSD, Alpha= .01, $CI_{.95}$= +/- .001

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100
3	**.0103**	**.0095**	**.0101**	**.0090**	**.0093**	**.0094**	**.0094**
4	**.0105**	**.0096**	**.0094**	**.0103**	**.0097**	.0088 ↓	.0087 ↓
5	**.0105**	.0083 ↓	.0083 ↓	**.0091**	.0083 ↓	.0087 ↓	.0087 ↓
6	**.0102**	**.0092**	**.0095**	.0075 ↓	.0088 ↓	.0077 ↓	.0085 ↓
8	**.0095**	.0082 ↓	.0082 ↓	.0078 ↓	.0085 ↓	.0073 ↓	.0075 ↓
10	**.0092**	.0078 ↓	.0079 ↓	.0080 ↓	.0070 ↓	.0069 ↓	.0075 ↓

↓ indicates p below $CI_{.95}$ **BOLD** indicates p within $CI_{.95}$ ↑ indicates p above $CI_{.95}$

Mean	SD	Min.	Max.	% ↓	% within	% ↑
.008772	.000977	.0069	.0105	57.1	42.9	0

Table 9. Differences in Per-Experiment and Experimentwise Type I Errors for HSD, Alpha= .01

K	n= 5	n= 10	n= 15	n= 20	n= 30	n= 60	n=100	Mean
3	.0014	.0010	.0009	.0010	.0007	.0008	.0009	.0010
4	.0026	.0013	.0015	.0018	.0013	.0012	.0012	.0016
5	.0028	.0014	.0016	.0018	.0012	.0014	.0013	.0016
6	.0028	.0022	.0023	.0011	.0019	.0015	.0015	.0019
8	.0030	.0023	.0022	.0018	.0025	.0016	.0015	.0021
10	.0033	.0025	.0024	.0025	.0020	.0017	.0021	.0024
Mean	.0027	.0018	.0018	.0017	.0016	.0014	.0014	.0018

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029421

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THE TUKEY HONESTLY SIGNIFICANT DIFFERENCE PROCEDURE AND ITS CONTROL OF THE TYPE I ERROR-RATE

Author(s): J. JACKSON BARNETTE & JAMES E. McLEAN

Corporate Source:

Publication Date: 11/5/98

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]

Printed Name/Position/Title: J. JACKSON BARNETTE ASSOC. PROF.

Organization/Address: UNIVERSITY OF IOWA 1-204 MEB IOWA CITY, IA 52242

Telephone: (319) 335 8905

FAX: (319) 335 8904

E-Mail Address: JACK-BARNETTE@UIOWA.EDU

Date: 12/10/98